WHY DO NONINVERTIBLE ESTIMATED MOVING AVERAGES OCCUR?

TECHNICAL REPORT NO. 13

T. W. ANDERSON AND AKIMICHI TAKEMURA

NOVEMBER 1984

U. S. ARMY RESEARCH OFFICE
CONTRACT DAAG29-82-K-0156
THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

WHY DO NONINVERTIBLE ESTIMATED MOVING AVERAGES OCCUR?

TECHNICAL REPORT NO. 13

T. W. ANDERSON
STANFORD UNIVERSITY

and

AKIMICHI TAKEMURA
UNIVERISTY OF TOKYO

NOVEMBER 1984

U. S. ARMY RESEARCH OFFICE
CONTRACT DAAG29-82-K-0156

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

WHY DO NONINVERTIBLE ESTIMATED MOVING AVERAGES OCCUR?

by

T. W. Anderson
Stanford University

and

Akimichi Takemura
University of Tokyo

ABSTRACT

The positive probability that an estimated moving average process is noninvertible is studied for maximum likelihood estimation of a univariate process. Upper and lower bounds for the probability in the first-order case are obtained as well as limits when the sample size tends to infinity. Higher order moving average models and autoregressive moving average models are also treated.

Key words: moving average models, maximum likelihood estimation, noninvertible moving average, autoregressive moving average processes.

WHY DO NONINVERTIBLE ESTIMATED MOVING AVERAGES OCCUR?

T. W. Anderson* and Akimichi Takemura
Stanford University · University of Tokyo

1. INTRODUCTION

In maximum likelihood estimation of a moving average process noninvertible estimates frequently appear, both with actual data and in simulation studies. Kang (1975) showed how noninvertibility occurs in the moving average of order one and indicated why it should be expected with positive probability. Cryer and Ledolter (1981) and Davidson (1981a), (1981b) have investigated the probabilities in finite samples. Sargan and Bhargava (1983) and Pesaran (1983) have found the limit of the probability that a noninvertible value is a local maximum of the likelihood function when the true value is noninvertible. We develop, organize, and generalize these results. Some new theoretical results include: i) a rigorous derivation of the limiting probabilities that the likelihood function attains a local maximum at a noninvertible value for noninvertible (Theorem 4.1) and invertible (Theorem 4.2) processes, ii) a lower bound for the probability that the likelihood function attains a global maximum at a noninvertible value (Corollary 5.1), iii) relations between maximum likelihood estimation and several least

*The research of the first author was carried out in part as Wesley Claire Mitchell Visiting Professor of Economics at Columbia University and at IBM Systems Research Institute.

square estimations (Theorem 5.1). Some similar results are obtained for the moving average process of general order (Section 6) and autoregressive and moving average processes of general order (Section 8). It will be shown that in general the region of possible autocovariances for a time series of finite length T is larger than the region corresponding to moving average processes $(T=\infty)$. This results in a positive probability that the estimated process falls on the boundary of the region of moving average processes, namely, the estimated process is noninvertible. This point will be clarified by studying the Jacobian matrix associated with the transformation from moving average coefficients to autocovariances (Theorem 6.1, Theorem 6.2, and Theorem 6.3) and by interpreting the results from a geometric viewpoint (Theorem 6.4). The above general consideration is illustrated by the MA(2) process (Section 7) and the ARMA (1,1) process (Section 9).

2. THE MOVING AVERAGE PROCESS OF ORDER ONE

The topic we are studying can be indicated by the simplest case, namely, the moving average of order one, designated as MA(1). Let $\{y_t\}$ be a stochastic process defined by

$$y_t = v_t + \alpha v_{t-1}, \quad t = \ldots,-1,0,1,\ldots, \tag{1}$$

where $\{v_t\}$ is a sequence of unobservable random variables with the properties

(2) $$\mathcal{E} v_t = 0 , \ \mathcal{E} v_t^2 = \sigma_v^2 , \ \mathcal{E} v_t v_s = 0 , \ t \neq s \ .$$

Then $\{y_t\}$ is a stochastic process stationary in the wide sense. If the v_t's are independent and identically distributed, $\{y_t\}$ is strictly stationary.

If $|\alpha| < 1$, we can invert (1) to obtain the autoregressive representation (of infinite order)

(3) $$v_t = y_t - \alpha y_{t-1} + \alpha^2 y_{t-2} + \dots \ .$$

The process is noninvertible if $\alpha = \pm 1$. Then (3) will not converge and the expression is meaningless.

The autoregressive representation is important because it is used for prediction. The prediction of y_t from $y_{t-1}, y_{t-2}, \dots$ is

$$\hat{y}_t = \mathcal{E} \ y_t | y_{t-1}, \ \dots = \alpha \ y_{t-1} - \alpha^2 y_{t-2} + \dots \ ,$$

known as exponential smoothing. Another reason for concern about invertibility is that iterative computational procedures may not converge if the estimate is ± 1 . Moreover, an appeal of the MA(1) model is that it approximates an autoregressive model with coefficients decreasing roughly exponentially.

The first and second-order moments of the observable process $\{y_t\}$ can be obtained from (1) and (2). The mean is

(4) $$\mathcal{E} \ y_t = 0 \ .$$

The autocovariance sequence is

$$(5) \qquad \mathcal{E}\, y_t^2 = \mathcal{E}(v_t + \alpha v_{t-1})^2 = \sigma_v^2(1+\alpha^2) = \sigma(0) ,$$

$$\mathcal{E}\, y_t y_{t-1} = \mathcal{E}(v_t + \alpha v_{t-1})(v_{t-1} + \alpha v_{t-2}) = \sigma_v^2 \alpha = \sigma(1) ,$$

$$\mathcal{E}\, y_t y_{t-h} = 0 = \sigma(h) , \quad h = 2,3,\ldots,$$

and $\sigma(-h) = \sigma(h)$. The first-order autocorrelation is

$$(6) \qquad \rho = \frac{\sigma(1)}{\sqrt{\sigma(0)}\,\sqrt{\sigma(0)}} = \frac{\sigma_v^2 \alpha}{\sigma_v^2(1+\alpha^2)} = \frac{\alpha}{1+\alpha^2} .$$

The other autocorrelations are 0. If α is replaced by its reciprocal

$$(7) \qquad \rho = \frac{1/\alpha}{1+(1/\alpha)^2} = \frac{\alpha}{1+\alpha^2} ;$$

the autocorrelation is unchanged. We can, therefore, restrict α to $-1 \le \alpha \le 1$ without loss of generality as far as first and second-order moments are concerned.

We shall assume $\{y_t\}$ is Gaussian, that is, all joint distributions are normal. Then the moments (4) and (5) completely describe the process. We note that ρ as a function of α is monotonically increasing in the interval [-1,1] and satisfies the inequality

$$(8) \qquad -\frac{1}{2} \le \rho \le \frac{1}{2} .$$

3. MAXIMUM LIKELIHOOD ESTIMATION FOR THE FIRST-ORDER MOVING AVERAGE

The observations on $\{y_t\}$ at T successive time points constitute a vector

$$y = (y_1,\ldots,y_T)' \quad . \tag{8}$$

It is an observation from a normal distribution with mean 0 and covariance matrix

$$\Sigma = \sigma(0)\, R \quad , \tag{9}$$

where

$$R = I_T + 2\rho A \quad , \tag{10}$$

and

$$A = \frac{1}{2}\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} . \tag{11}$$

Then the logarithm of the likelihood function is

$$\log L = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log \sigma(0) - \frac{1}{2}\log|\underset{\sim}{R}| - \frac{1}{2\sigma(0)}\underset{\sim}{y}'\underset{\sim}{R}^{-1}\underset{\sim}{y} \ .$$

For given ρ the value of $\sigma(0)$ that maximizes the likelihood function is

$$\hat{\sigma}(0) = \frac{\underset{\sim}{y}'\underset{\sim}{R}^{-1}\underset{\sim}{y}}{T} \ . \tag{13}$$

Then the logarithm of the concentrated likelihood function is (except for constants)

$$M(\rho) = -\log|\underset{\sim}{R}| - T\log \underset{\sim}{y}'\underset{\sim}{R}^{-1}\underset{\sim}{y} \ . \tag{14}$$

If $\underset{\sim}{y} \neq \underset{\sim}{0}$, the maximum of $M[\rho(\alpha)]$ with respect to α exists [Anderson and Mentz (1980)]. The derivative equation is

$$\frac{dM}{d\alpha} = \frac{dM}{d\rho}\frac{d\rho}{d\alpha} = \frac{dM}{d\rho}\frac{1-\alpha^2}{(1+\alpha^2)^2} \ . \tag{15}$$

The derivative is 0 at $\alpha = \pm 1$ [Kang (1975)]. The question is when is $\alpha = 1$ or -1 a maximum?

If $\underset{\sim}{y} \neq \underset{\sim}{0}$, the maximum of $M(\rho)$ exists such that $\underset{\sim}{R}$ is positive definite. It will be shown that $\underset{\sim}{R}$ is positive definite for $-a < \rho < a$, where $\frac{1}{2} < a < 1$. If $dM(\rho)/d\rho \neq 0$ for all ρ in the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$, then $\alpha = 1$ or -1 yields a maximum. The maximum can be at $\alpha = 1$ or -1 if $dM(\rho)/d\rho = 0$ for some values of ρ in the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$.

A local maximum occurs at $\alpha = 1$ if and only if

$$(16)\qquad 0 > \left.\frac{d^2M[\rho(\alpha)]}{d\alpha^2}\right|_{\alpha=1} = \frac{d}{d\alpha}\left[\frac{dM}{d\rho}\,\frac{d\rho}{d\alpha}\right]_{\alpha=1}$$

$$= \left.\frac{d^2M}{d\rho^2}\left(\frac{d\rho}{d\alpha}\right)^2_{\alpha=1} + \frac{dM}{d\rho}\,\frac{d^2\rho}{d\alpha^2}\right|_{\alpha=1}$$

$$= -\frac{1}{2}\,\frac{dM}{d\rho}$$

because $d\rho/d\alpha = 0$ at $\alpha = 1$ and $d^2\rho/d\alpha^2 = -\frac{1}{2}$ at $\alpha = 1$. Thus a local maximum occurs at $\alpha = 1$ if and only if

$$(17)\qquad \left.\frac{dM}{d\rho}\right|_{\rho=\frac{1}{2}} > 0 \quad .$$

To study the probabilities of maxima it is convenient to put the concentrated likelihood in a canonical form. Let $\underset{\sim}{P} = (p_{ij})$, where

$$(18)\qquad p_{st} = \sqrt{\frac{2}{T+1}}\,\sin\frac{\pi\, st}{T+1}\ , \qquad s,t = 1,\dots,T \quad .$$

Then $\underset{\sim}{P}'\underset{\sim}{A}\underset{\sim}{P} = \underset{\sim}{D}$ is diagonal and the diagonal elements are

$$(19)\qquad d_t = \cos\frac{\pi t}{T+1}\ , \qquad t = 1,\dots,T$$

[Anderson (1971), Sec. 6.5]. The roots can be visualized by dividing one-half of the unit circle into $T+1$ equal parts and projecting the points on the circumference on the diameter. Then

$$(20)\qquad \underset{\sim}{P}'\underset{\sim}{R}\underset{\sim}{P} = \underset{\sim}{I}_T + 2\rho\underset{\sim}{D}\ ,$$

(21) $$|\underset{\sim}{R}| = \prod_{t=1}^{T} (1+2\,\rho d_t) \quad ,$$

(22) $$\underset{\sim}{y}'\underset{\sim}{R}^{-1}\underset{\sim}{y} = \sum_{t=1}^{T} \frac{z_t^2}{1+2\,\rho d_t} \quad ,$$

where

(23) $$\underset{\sim}{y} = \underset{\sim}{P}\,\underset{\sim}{z} \quad .$$

Since $\underset{\sim}{y}$ has the distribution $N[\underset{\sim}{0}, \sigma(0)\underset{\sim}{R}]$, then $\underset{\sim}{z}$ has the distribution $N[\underset{\sim}{0}, \sigma(0)(\underset{\sim}{I}_T + 2\,\rho \underset{\sim}{D})]$. The logarithm of the concentrated likelihood function is (except for constants)

$$M(\rho) = -\sum_{t=1}^{T} \log(1+2\,\rho d_t) - T \log\left[\sum_{t=1}^{T} \frac{z_t^2}{1+2\,\rho d_t}\right] \quad .$$

For $\underset{\sim}{R}$ to be positive definite, $\underset{\sim}{I}_T + 2\,\rho \underset{\sim}{D}$ must be positive definite. Thus $1 + 2\,\rho d_t > 0$, $t=1,\ldots,T$. These imply

(25) $$-\frac{1}{2\cos\frac{\pi}{T+1}} < \rho < \frac{1}{2\cos\frac{\pi}{T+1}} \quad .$$

The maximum likelihood estimator of ρ is a solution of the derivative of the logarithm of the likelihood function set equal to 0 . The derivative is

(26) $$\frac{dM}{d\rho} = -\sum_{t=1}^{T} \frac{2d_t}{1+2\,\rho d_t} + \frac{T}{\sum_{t=1}^{T} \frac{z_t^2}{1+2\,\rho\, d_t}} \sum_{t=1}^{T} \frac{2d_t z_t^2}{(1+2\,\rho d_t)^2} \quad .$$

The likelihood equation is

$$(27)\qquad T\sum_{t=1}^{T}\frac{2d_t z_t^2}{(1+2\,\rho d_t)^2} - \sum_t \frac{2d_t}{1+2\,\rho d_t}\sum_t \frac{z_t^2}{1+2\,\rho d_t} = 0 \; .$$

The left-hand side is a polynomial in ρ of degree $2T-3$ or less.

4. A LOCAL MAXIMUM AT A NONINVERTIBLE VALUE

The probability of a local maximum at $\alpha = 1$ is

$$(28)\qquad \Pr\left\{\left.\frac{dM}{d\rho}\right|_{\rho=\frac{1}{2}} > 0\right\} = \Pr\left\{\sum_{t=1}^{T}\frac{(T+2)d_t + T - 1}{(1+d_t)^2}\; z_t^2 > 0\right\} .$$

We have made use of the fact (proved in A1 in the appendix) that

$$(29)\qquad \sum_{t=1}^{T}\frac{d_t}{1+d_t} = -\frac{T(T-1)}{3} \; .$$

If we let

$$(30)\qquad z_t = \sqrt{1+2\,\rho d_t}\;\; x_t \; ,$$

then x_t has the distribution $N(0,1)$. We can write (28) as

$$(31)\qquad \Pr\left\{\sum_{t=1}^{T}\frac{(T+2)d_t + (T-1)}{(1+d_t)^2}\;(1+2\,\rho d_t)x_t^2 > 0\right\} .$$

As a simple case consider $T=2$. The concentrated likelihood function is (except for a constant)

(32) $$e^{\frac{1}{2}M(\rho)} = \frac{1}{(1-\rho^2)^{\frac{1}{2}}\left[\frac{z_1^2}{1+\rho} + \frac{z_2^2}{1-\rho}\right]}$$

$$= \frac{1}{\frac{1}{\gamma}z_1^2 + \gamma\, z_2^2} \quad ,$$

where

(33) $$\gamma = \sqrt{\frac{1+\rho}{1-\rho}} \quad .$$

Then γ is a monotonically increasing function of ρ. A small table of this function is

Table 1

ρ	γ
-1	0
$-\frac{1}{2}$	$1/\sqrt{3}$
0	1
$\frac{1}{2}$	$\sqrt{3}$
1	∞

The likelihood function is unimodal (Figure 1). The unique maximum is at

(34) $$\hat{\gamma} = \frac{|z_1|}{|z_2|} \quad .$$

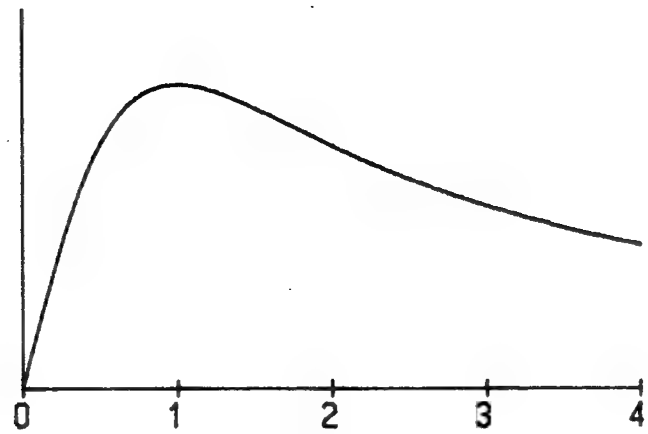

Figure 1. Likelihood function for MA(1), T = 2

Thus the probability of the maximum at $\alpha = 1$ is

$$
\begin{aligned}
(35) \qquad \Pr\{\hat{\alpha} = 1\} &= \Pr\{\hat{\gamma} > \sqrt{3}\} \\
&= \Pr\left\{\frac{x_1^2}{x_2^2} > 3\,\frac{1-\rho}{1+\rho}\right\} \\
&= 1 - \frac{2}{\pi}\,\text{arc tan}\sqrt{3\,\frac{1-\rho}{1+\rho}} \\
&= \begin{cases} \frac{1}{3}, & \rho = 0\,, \\ \frac{1}{2}, & \rho = \frac{1}{2}\,. \end{cases}
\end{aligned}
$$

The probability (31) has been evaluated for different values of ρ (or α) by several numerical techniques by Pesaran (1983) and by Cryer and Ledolter (1981). The Table 2 of probabilities of local maxima was

given by the latter. Because of symmetry, the probability that $\alpha = -1$ is a local maximum for a process coefficient α is the probability that $\alpha = 1$ is a local maximum for a process parameter $-\alpha$.

Table 2

Probabilities of Local Maxima at 1 and -1

α	T = 2 local max		T = 10 local max		T = 25 local max	
	1	-1	1	-1	1	-1
0.0	.333	.333	.102	.102	.015	.015
.2	.389	.282	.145	.076	.025	.011
.4	.440	.244	.223	.062	.050	.009
.6	.476	.220	.353	.056	.119	.008
.8	.495	.208	.533	.053	.319	.008
1.0	.500	.205	.637	.053	.649	.007

Now we consider large sample size. First we find the limiting probability of a relative maximum at $\alpha = 1$ when $\rho = \frac{1}{2}(\alpha = 1)$.

Theorem 4.1:

$$\lim_{T\to\infty} \Pr\left\{ \sum_{t=1}^{T} \frac{(T+2)d_t + T - 1}{1+d_t} x_t^2 > 0 \right\} = \Pr\{W^2 \le \tfrac{1}{6}\}$$

$$= .6575 ,$$

where W^2 is the limiting form of the Cramér-von Mises statistic.

Proof: For $\rho = \frac{1}{2}$ (31) is

(36)
$$\Pr\left\{\sum_{t=1}^{T} x_t^2 - \frac{3}{T+2}\sum_{t=1}^{T}\frac{x_t^2}{1+d_t} > 0\right\}$$
$$= \Pr\left\{\frac{3}{T(T+2)}\sum_{t=1}^{T}\frac{x_t^2}{1+d_t} < \frac{1}{T}\sum_{t=1}^{T} x_t^2\right\} .$$

Let K_T be a sequence of integers such that $K_T \to \infty$ as $T \to \infty$, $K_T/T \to 0$, and $K_T^2/T \to \infty$. Then write (36) as

(37)
$$\Pr\left\{\sum_{t=T-K_T}^{T}\left(\frac{3}{T(T+2)(1+d_t)} - \frac{1}{T}\right)x_t^2 < \sum_{t=1}^{T-K_T-1}\left[\frac{1}{T} - \frac{3}{T(T+2)(1+d_t)}\right]x_t^2\right\} .$$

The two sides of the inequality in (37) are independent. For $k = 0, 1, \ldots, K_T$

(38)
$$1 + d_{T-k} = 1 + \cos\frac{\pi(T-k)}{T+1} = 1 - \cos\frac{\pi(k+1)}{T+1}$$
$$= 1 - \left\{1 - \frac{1}{2}\left[\frac{\pi(k+1)}{T+1}\right]^2 + \mathcal{O}\left[\left(\frac{K_T}{T}\right)^4\right]\right\}$$
$$= \frac{1}{2}\frac{\pi^2(k+1)^2}{(T+1)^2} + \mathcal{O}\left[\left(\frac{K_T}{T}\right)^4\right] .$$

Then the coefficient of x_{T-k}^2 is

$$(39)\qquad \frac{3}{T(T+2)(1+d_{T-k})} - \frac{1}{T} = \frac{6(T+1)^2}{T(T+2)} \frac{1}{\pi^2(k+1)^2} - \frac{1}{T} + \mathcal{O}(T^{-2})$$

$$= \frac{6}{\pi^2(k+1)^2} + \mathcal{O}(T^{-1}) \quad .$$

The left-hand side of the inequality in (37) has the limiting distribution of

$$(40)\qquad 6 \sum_{j=1}^{\infty} \frac{x_j'^2}{\pi^2 j^2} = 6\, W^2 \;,$$

where W^2 is the limiting form of the Cramér-von Mises statistic and $\{x_j'\}$ are independent $N(0,1)$ variables.

Since $\cos\theta$ is a decreasing function of $\theta (0<\theta<\pi)$, for $t=1,\ldots,T-K_T-1$

$$(41)\qquad \frac{3}{T(T+2)(1+d_t)} < \frac{3}{T(T+2)\left[1+\cos\dfrac{\pi(T-K_T)}{T+1}\right]}$$

$$\sim \frac{6}{\pi^2 K_T^2} \quad .$$

The right-hand side of the inequality in (37) is

$$(42)\qquad \frac{1}{T}\sum_{t=1}^{T-K_T-1} x_t^2 + \mathcal{O}_p\left(\frac{T}{K_T^2}\right) = \left(1 - \frac{K_T+1}{T}\right)\frac{1}{T-K_T-1}\sum_{t=1}^{T-K_T-1} x_t^2 + \mathcal{O}_p\left(\frac{T}{K_T^2}\right)$$

which has 1 as a probability limit. Hence the limit of (37) is $\Pr\{W^2 \le 1/6\}$. The value of this by interpolation in the table of the distribution function given by Anderson and Darling (1952) is .6575. Q.E.D.

Sargan and Bhargava (1983) have obtained this result by a different method, Pesaran (1983) has used a somewhat similar method.

We shall now show that if the MA(1) process is invertible, the probability of a local maximum at $\alpha = 1$ or $\alpha = -1$ goes to 0 at least as fast as T^{-n} for any n.

Theorem 4.2. Let $-\frac{1}{2} < \rho < \frac{1}{2}$ be fixed. Then for any n

$$\lim_{T\to\infty} T^n \Pr\left\{\sum_{t=1}^{T} \frac{(T+2)d_t + T-1}{(1+d_t)^2} (1+2\rho d_t)\, x_t^2 > 0\right\} = 0 . \tag{43}$$

Proof: Let

$$w_t = -\frac{(T+2)d_t + T-1}{(1+d_t)^2} = \frac{3}{(1+d_t)^2} - \frac{T+2}{1+d_t} . \tag{44}$$

Then the probability on the left-hand side of (43) is

$$\Pr(\Sigma_{t=1}^{T} w_t(1+2\rho d_t)\, x_t^2 < 0) . \tag{45}$$

We first investigate the behavior of w_t. Let

$$f(x) = \frac{3}{(1+x)^2} - \frac{T+2}{1+x} = -\frac{(T+2)x + T-1}{(1+x)^2} , \qquad x > -1 . \tag{46}$$

As $x \to -1$, $f(x) \to \infty$. We have

$$
(47) \qquad f(x) \begin{cases} > 0 & \text{if} \quad x < -\dfrac{T-1}{T+2}, \\ = 0 & \text{if} \quad x = -\dfrac{T-1}{T+2}, \\ < 0 & \text{if} \quad x > -\dfrac{T-1}{T+2}, \end{cases}
$$

and

$$
\begin{aligned}
(48) \qquad f'(x) &= -\frac{6}{(1+x)^3} + \frac{T+2}{(1+x)^2} \\
&= \frac{1}{(1+x)^3}\{(T+2)(1+x)-6\} \\
&= \frac{1}{(1+x)^3}\{x(T+2)+(T-4)\}.
\end{aligned}
$$

This shows that f is decreasing to the left of $-(T-4)/(T+2)$ and increasing to the right of $-(T-4)/(T+2)$. See Figure 2.

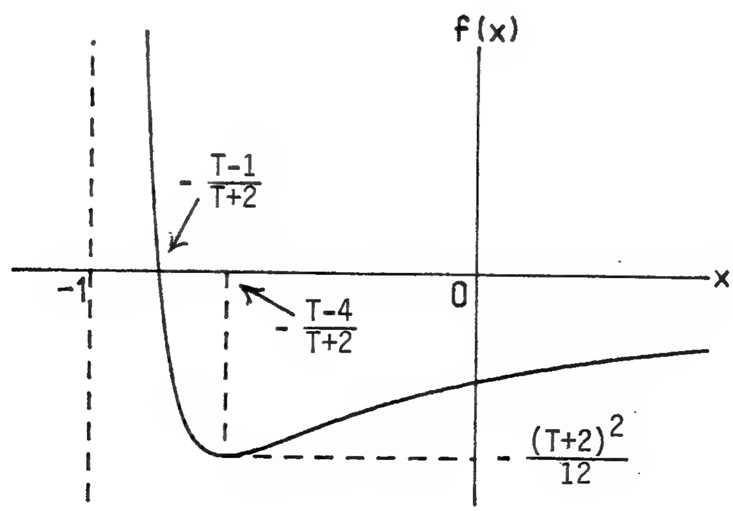

Figure 2.

Also note that the minimum of f is

(49)
$$f\left(-\frac{T-4}{T+2}\right) = -\frac{-(T+2)\frac{T-4}{T+2} + T - 1}{\left(1-\frac{T-4}{T+2}\right)^2}$$
$$= \frac{T-4-(T-1)}{6^2}(T+2)^2$$
$$= -\frac{(T+2)^2}{12} \quad .$$

Separating positive and negative weights we write (45) as

(50)
$$\Pr\left\{\sum_{w_t>0} w_t(1+2\rho d_t)x_t^2 < \sum_{w_t\le 0}\left[-w_t(1+2\rho d_t)x_t^2\right]\right\} \; .$$

To bound this probability from above we try to make the left-hand side smaller and right-hand side larger. Note that now the weights are all positive.

Right-hand side. We saw that $w_t \ge -\frac{(T+2)^2}{12}$. Hence $-w_t \le \frac{(T+2)^2}{12}$. Also

(51)
$$1+2\rho d_t \le 1+2|\rho| < 2 \; .$$

Hence the right-hand side is bounded from above by

(52)
$$\frac{(T+2)^2}{6}\sum_{w_t\le 0} x_t^2 \quad .$$

<u>Left-hand side</u>. w_t is positive for $t = T, T-1, \ldots$. By taking only a finite number k of terms on the left-hand side we decrease it; that is ,

$$\sum_{w_t>0} w_t(1+2\rho d_t)x_t^2 \geq \sum_{j=0}^{k-1} w_{T-j}\ (1+2\rho d_{T-j})x_{T-j}^2 \quad . \tag{53}$$

(See A2 in the Appendix for verification that the left-hand side of (53) contains at least k terms for large T .)

Now we look at the weights

$$w_{T-j}(1+2\rho d_{T-j})\ , \qquad j = 0,\ldots,k-1 \quad . \tag{54}$$

w_{T-j} is decreasing in j ; hence

$$w_{T-j} \geq w_{T-k+1}\ , \qquad j = 0,\ldots,k-1 \quad . \tag{55}$$

Also

$$1+2\rho d_{T-j} \geq 1 - 2|\rho| \quad . \tag{56}$$

Hence

$$\sum_{j=0}^{k-1} w_{T-j}(1+2\rho d_{T-j})x_{T-j}^2 \ \geq\ w_{T-k+1}(1-2|\rho|) \sum_{j=0}^{k-1} x_{T-j}^2 \quad . \tag{57}$$

Combining these results we find that (50) is less than or equal to

$$\Pr\left\{w_{T-k+1}(1-2|\rho|)\chi_I^2 \le \frac{(T+2)^2}{6}\,\chi_{II}^2\right\} , \tag{58}$$

where χ_I^2 and χ_{II}^2 are independent chi-square random variables with k and T' (= number of $w_t > 0$) degrees of freedom. Hence, (58) is

$$E_{\chi_{II}^2}\left[c_1 \int_0^{c_T\chi_{II}^2} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}\,dx\right] \tag{59}$$

$$\le c_1\, E_{\chi_{II}^2}\left[\int_0^{c_T\chi_{II}^2} x^{\frac{k}{2}-1}\,dx\right]$$

$$= c_2\, E_{\chi_{II}^2}\left[(c_T\,\chi_{II}^2)^{\frac{k}{2}}\right] ,$$

where $c_1 = [2^{k/2}\,\Gamma(k/2)]^{-1}$, $c_2 = 2c_1/k$, and

$$c_T = \frac{(T+2)^2}{6} \cdot \frac{1}{(1-2|\rho|)w_{T-k+1}} . \tag{60}$$

Then for k even the right-hand side of (59) is

$$c_2\, c_T^{k/2}\, T'(T'+2)\ldots(T'+k-2) . \tag{61}$$

There are $k/2$ terms in the product, which is $\mathcal{O}(T^{k/2})$ because $T' = \mathcal{O}(T)$. However

(62)
$$w_{T-k+1} = \frac{3}{(1+d_{T-k+1})^2} - \frac{T+2}{1+d_{T-k+1}}$$
$$= \Theta(T^4) .$$

Hence $c_T = \Theta(T^{-2})$, and $c_T^{k/2} = \Theta(T^{-k})$. Combining these we have (50) is less than or equal to $\Theta(T^{-k/2})$. This completes the proof. Q.E.D.

5. LEAST SQUARES ESTIMATION FOR THE FIRST-ORDER MOVING AVERAGE

We now consider two kinds of least squares estimators and investigate the relation between these and the maximum likelihood estimator. As a corollary we shall obtain a lower bound for the probability that the likelihood function attains a global maximum at $\alpha = 1$. There are two ways of parametrizing the process. The parametrization $(\sigma(0),\rho)$ has been used above. Another parametrization is (σ_v^2,α) where $\sigma_v^2 = \mathcal{E}v_t^2$ is the variance of the disturbance term. Let $\underset{\sim}{Q} = (1+\alpha^2)\underset{\sim}{I}_T + 2\alpha\underset{\sim}{A}$. Then the logarithm of the likelihood function is

(63)
$$\log L = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log \sigma_v^2 - \frac{1}{2}\log|\underset{\sim}{Q}| - \frac{1}{2\sigma_v^2}\underset{\sim}{y}'\underset{\sim}{Q}^{-1}\underset{\sim}{y} .$$

Maximizing with respect to σ_v^2, we obtain the concentrated likelihood function

(64)
$$M_{II}(\alpha) = -\log|\underset{\sim}{Q}| - T\log \underset{\sim}{y}'\underset{\sim}{Q}^{-1}\underset{\sim}{y} .$$

Ignoring the determinant term,consider two quadratic forms

(65.) $$S_I(\rho) = \underset{\sim}{y}'\underset{\sim}{R}^{-1}\underset{\sim}{y} \quad ,$$

(66) $$S_{II}(\alpha) = \underset{\sim}{y}'\underset{\sim}{Q}^{-1}\underset{\sim}{y} \quad .$$

Let $\hat{\alpha}_{LS,I}$ and $\hat{\alpha}_{LS,II}$ denote estimators which minimize $S_I[\rho(\alpha)]$ and $S_{II}(\alpha)$ respectively:

(67) $$\min_\alpha S_I[\rho(\alpha)] = \min_\alpha S_I[\alpha/(1+\alpha^2)] \text{ at } \hat{\alpha}_{LS,I} \, ,$$

$$\min_\alpha S_{II}(\alpha) \text{ at } \hat{\alpha}_{LS,II} \, .$$

Furthermore,let $\hat{\alpha}_{ML}$ denote the maximum likelihood estimator.

Theorem 5.1.

(68) $$|\hat{\alpha}_{LS,I}| \le |\hat{\alpha}_{ML}| \le |\hat{\alpha}_{LS,II}| \, ,$$

and

(69) $$\hat{\alpha}_{LS,I} = 1 \Rightarrow \hat{\alpha}_{ML} = 1 \Rightarrow \hat{\alpha}_{LS,II} = 1 \quad .$$

Proof. Consider $\log|\underset{\sim}{R}| = \Sigma_{t=1}^{T} \log(1+2\rho d_t) = \Sigma_{t=1}^{[T/2]} \log(1-4\rho^2 d_t^2)$. Clearly $\log|\underset{\sim}{R}|$ is strictly decreasing in ρ^2 . Let $\hat{\rho}_{LS,I} = \hat{\alpha}_{LS,I}/(1+\hat{\alpha}_{LS,I}^2)$. Then $S_I(\hat{\rho}_{LS,I}) \le S_I(\rho)$ for all ρ . Hence for all $|\rho|<|\hat{\rho}_{LS,I}|$

we have $M(\rho) - M(\hat{\rho}_{LS,I}) = (\log|\underset{\sim}{R}|_{\rho=\hat{\rho}_{LS,I}} - \log|\underset{\sim}{R}|) - T[\log S_I(\rho) - \log S_I(\hat{\rho}_{LS,I})] < 0$. This implies that $|\hat{\alpha}_{ML}| \geq |\hat{\alpha}_{LS,I}|$. To prove the second inequality, consider $\log|\underset{\sim}{Q}|$. Since $|\underset{\sim}{Q}| = 1 + \alpha^2 + \ldots + \alpha^{2T}$, $\log\ |\underset{\sim}{Q}|$ is strictly increasing in α^2 . Hence a similar argument as above yields $|\hat{\alpha}_{ML}| \leq |\hat{\alpha}_{LS,II}|$. This proves (68). (69) follows from (68) by virtue of the fact $0 = \Pr[\ S_I(\frac{1}{2}) = S_I(-\frac{1}{2})] = \Pr[M(\frac{1}{2}) = M(-\frac{1}{2})] = \Pr[S_{II}(1) = S_{II}(-1)]$.

Corollary 5.1.

(70) $$\Pr(\hat{\alpha}_{ML} = 1) \geq \Pr(\hat{\alpha}_{LS,I} = 1) = \Pr\left\{\sum_t \frac{d_t}{(1+d_t)^2}\ z_t^2 \geq 0\right\} .$$

Proof: The first inequality is an immediate consequence of (69). Now the event $\hat{\alpha}_{LS,I} = 1$ is equivalent to $S_I(\rho) \geq S_I(\frac{1}{2})$, which is

(71) $$\sum_t \frac{z_t^2}{1+2\rho d_t} \geq \sum_t \frac{z_t^2}{1+d_t} , \qquad \forall\ \rho \in \left[-\frac{1}{2}, \frac{1}{2}\right] ,$$

or

(72) $$\sum_t \frac{d_t}{(1+2\rho d_t)(1+d_t)}\ z_t^2 \geq 0 , \qquad \forall\ \rho \in \left[-\frac{1}{2}, \frac{1}{2}\right] .$$

The event is

(73) $$\min_\rho \sum_t \frac{d_t}{(1+2\rho d_t)(1+d_t)}\ z_t^2 \geq 0 .$$

Since each coefficient in (73) is decreasing in ρ , the event is

$$\sum_t \frac{d_t}{(1+d_t)^2} z_t^2 \geq 0 \quad . \tag{74}$$

Q.E.D.

When $T=2$, the lower bound is

$$\Pr\left\{ \frac{z_1^2}{z_2^2} \geq 9 \middle| \rho \right\} = \Pr\left\{ \frac{x_1^2}{x_2^2} \geq 9 \frac{1-\rho}{1+\rho} \right\} \tag{75}$$

$$= 1 - \frac{2}{\pi} \text{arc tan } 3\sqrt{\frac{1-\rho}{1+\rho}}$$

$$= \begin{cases} .204, & \rho=0 , \\ \frac{1}{3} , & \rho = \frac{1}{2} . \end{cases}$$

The lower bounds of .204 and .333 are to be compared with the exact values of .333 and .5, respectively. Analysis similar to that applied to the upper bound (i.e. local maximum) shows that for $\rho = \frac{1}{2}$ as $T \to \infty$ the lower bound approaches 0.

Theorem 5.2. $\hat{\alpha}_{LS,I}$ is biased toward origin. $\hat{\alpha}_{ML}$ and $\hat{\alpha}_{LS,II}$ are consistent.

It will be shown that if $\alpha = 1$, then $\hat{\alpha}_{LS,I}$ converges to .829 in probability. Because of this bias $\Pr(\hat{\alpha}_{LS,I} = 1 \mid \alpha = 1)$ goes to zero and this fact indicates that this lower bound is not sharp.

To prove the theorem we need the following lemma.

Lemma 1. Let $|a| \leq 1$, $|b| < 1$. Then

$$J(a;b) = \frac{1}{\pi} \int_0^\pi \frac{1 + a \cos \theta}{1 + b \cos \theta} d\theta \tag{76}$$

$$= \frac{a}{b} + \left(1 - \frac{a}{b}\right) \frac{1}{\sqrt{1-b^2}} .$$

Proof. Differentiating the relation

$$\int_0^\pi \log(a + b \cos \theta) d\theta = \pi \log \frac{1}{2} (a + \sqrt{a^2 - b^2}), \quad a > b \geq 0 \tag{77}$$

[Anderson (1971), Problem 69 of Chapter 6] with respect to a and setting $a = 1$ we obtain

$$\int_0^\pi \frac{1}{1 + b \cos \theta} d\theta = \frac{\pi}{\sqrt{1-b^2}} . \tag{78}$$

But $(1 + a \cos \theta)/(1 + b \cos \theta) = a/b + (1 - a/b)/(1 + b \cos \theta)$. The lemma follows. Q.E.D.

Let ρ^* be the true autocorrelation coefficient. Then

(79) $$S_I(\rho) = \sum_{t=1}^{T} \frac{z_t^2}{1+2\rho d_t} = \sum_{t=1}^{T} \frac{1+2\rho^* d_t}{1+2\rho d_t} x_t^2 ,$$

where the x_t's are independent standard normal variables. We consider $S_I(\rho)$ in the open interval $-\frac{1}{2} < \rho < \frac{1}{2}$. ρ^* can be equal to $\pm\frac{1}{2}$. Let ρ be fixed for the moment. Then as $T \to \infty$

(80) $$\begin{aligned} E_{\rho^*}[\tfrac{1}{T} S_I(\rho)] &= \frac{1}{T} \sum_{t=1}^{T} \frac{1+2\rho^* d_t}{1+2\rho d_t} \\ &\to \int_0^1 \frac{1+2\rho^* \cos(\pi s)}{1+2\rho \cos(\pi s)} ds \\ &= \frac{1}{\pi} \int_0^{\pi} \frac{1+2\rho^* \cos\theta}{1+2\rho \cos\theta} d\theta \\ &= J(2\rho^*;2\rho) . \end{aligned}$$

Furthermore

(81) $$\begin{aligned} \mathrm{Var}[\tfrac{1}{T} S_I(\rho)] &= \frac{1}{T^2} \sum_{t=1}^{T} 2\left(\frac{1+2\rho^* d_t}{1+2\rho d_t}\right)^2 \\ &\simeq \frac{2}{T} \int_0^1 \left[\frac{1+2\rho^* \cos(\pi s)}{1+2\rho \cos(\pi s)}\right]^2 ds \\ &\to 0 \end{aligned}$$

as $T \to \infty$. Hence for a fixed ρ, $\frac{1}{T} S_I(\rho)$ converges to $J(2\rho^*;2\rho)$ in probability. Now note that

$$\frac{\partial^2}{\partial b^2} J(a;b) = \frac{2}{\pi}\int_0^\pi \frac{\cos^2\theta(1+a\cos\theta)}{(1+b\cos\theta)^3}\,d\theta > 0 . \tag{82}$$

Hence for given a, $J(a;b)$ is convex in b and has a unique minimum, $b(a)$, say. By a standard argument then we see that the $\hat{\rho}_{LS,I}$ which minimizes $S_I(\rho)$ converges in probability to $\frac{1}{2}b(2\rho^*)$. An explicit expression for the minimum can be obtained by solving $\frac{\partial}{\partial b}J(a;b) = 0$. Now

$$\frac{\partial}{\partial b} J(a;b) = -\frac{a}{b^2} + \frac{a}{b^2}\frac{1}{\sqrt{1-b^2}} + \frac{b-a}{(1-b^2)^{3/2}} . \tag{83}$$

Solving this for a we obtain

$$a = \frac{b^3}{(1-b^2)^{3/2} + 2b^2 - 1} . \tag{84}$$

A plot of this relation in terms of $a/2 = \rho^*$ and $b/2 = \text{plim}\,\hat{\rho}_{LS,I}|\rho^*$ is given in Figure 3.

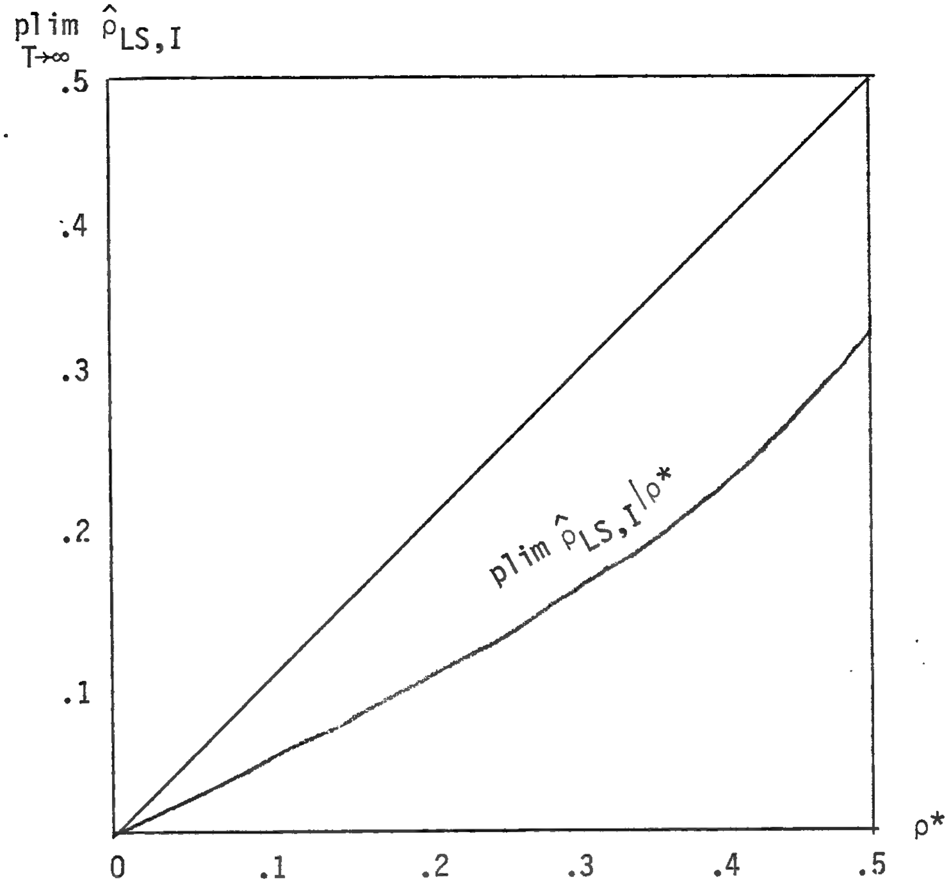

Figure 3.

This shows that $\hat{\rho}_{LS,I}$ is heavily downward biased. In particular if $\rho^* = \frac{1}{2}$, then plim $\hat{\rho}_{LS,I} = .354$ (or in terms of α, if $\alpha = 1$ then plim $\hat{\alpha}_{LS,I} = .829$).

As to $\hat{\alpha}_{LS,II}$ note that $S_{II}(\rho) = S_I(\rho(\alpha))/(1+\alpha^2)$ or

$$S_{II}[\alpha(\rho)] = \frac{1+\sqrt{1-4\rho^2}}{2} S_I(\rho) . \tag{85}$$

Let

(86) $$J_{II}(a;b) = \frac{1+\sqrt{1-b^2}}{2} J(a;b) = \frac{1}{2}[1+ \frac{1-ab}{\sqrt{1-b^2}}] .$$

Then

(87) $$\frac{\partial}{\partial b} J_{II}(a;b) = \frac{b-a}{(1-b^2)^{3/2}} .$$

Note that $\frac{\partial}{\partial b} J_{II}(a;b) < 0$ if $b < a$ and $\frac{\partial}{\partial b} J_{II}(a;b) > 0$ if $b > a$. Hence for given a $J_{II}(a;b)$ attains a unique minimum at $b = a$. By a standard argument again, then $\hat{\rho}_{LS,II}$ converges to ρ^* in probability showing that $\hat{\rho}_{LS,II}$ is consistent (and so is $\hat{\alpha}_{LS,II}$).

As to the maximum likelihood estimator consider $1/T$ times the concentrated likelihood function in terms of α:

(88) $$\frac{1}{T} M_{II}(\alpha) = -\frac{1}{T} \log|\underset{\sim}{Q}| - \log \underset{\sim}{y}'\underset{\sim}{Q}^{-1}\underset{\sim}{y}$$
$$= -\frac{1}{T} \log|\underset{\sim}{Q}| - \log S_{II}(\alpha) .$$

Now $|\underset{\sim}{Q}| = 1 + \alpha^2 + \ldots + \alpha^{2T}$, hence $1/T \log|\underset{\sim}{Q}| \le 1/T \log(T+1) \to 0$ as $T \to \infty$. Therefore the determinant term is asymptotically negligible. Therefore consistency of $\hat{\alpha}_{LS,II}$ implies consistency of $\hat{\alpha}_{ML}$. This proves the theorem. Q.E.D.

Note that $\log|\underset{\sim}{R}| = - T \log(1+\alpha^2) + \log|\underset{\sim}{Q}|$. Hence $\log|\underset{\sim}{R}|/T \to -\log(1+\alpha^2)$, which is not asymptotically negligible. This fact explains the inconsistency of $\hat{\alpha}_{LS,I}$.

As Theorem 5.1 shows, $\hat{\alpha}_{LS,II}$ is more likely to assume the noninvertible value ± 1 than the maximum likelihood estimator.

This can be also seen by considering the behavior of $S_{II}(\alpha)$ at $\alpha = \pm 1$. Note that

(89) $$S_{II}'(\alpha) = -2 \sum_{t=1}^{T} \frac{\alpha + d_t}{(1+\alpha^2+2\alpha d_t)^2} z_t^2 .$$

Hence

(90) $$S_{II}'(1) = -\frac{1}{2} \sum_{t=1}^{T} \frac{z_t^2}{1+d_t} < 0 .$$

Similarly $S_{II}'(-1) > 0$. This shows that $\alpha = \pm 1$ is always local minimum of the quadratic form $S_{II}(\alpha)$. Ignoring the determinant has a disadvantageous effect on estimation.

6. THE MOVING AVERAGE OF GENERAL ORDER

The results for the general order of moving average are not as clearcut. We write the process as

(91) $$y_t = \gamma_0 w_t + \gamma_1 w_{t-1} + \dots + \gamma_q w_{t-q} ,$$

where

(92) $$\gamma_0 > 0 , \qquad \mathcal{E} w_t = 0 , \qquad \mathcal{E} w_t^2 = 1 ,$$

and the w_t's are uncorrelated. Then the autocovariance sequence is

(93) $$\begin{aligned} \sigma(h) &= \sum_{j=0}^{q-h} \gamma_j \gamma_{j+h} , \quad h = 0,1,\dots,q , \\ &= 0 , \qquad h = q+1, \dots , \end{aligned}$$

and $\sigma(-h) = \sigma(h)$. The covariance matrix of T successive terms in the process is the $T \times T$ matrix

$$(94) \qquad \underset{\sim}{\Sigma}_T = \begin{bmatrix} \sigma(0) & \sigma(1) & \cdots & \sigma(q) & 0 & \cdots & 0 \\ \sigma(1) & \sigma(0) & \cdots & \sigma(q-1) & \sigma(q) & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \sigma(q) & \sigma(q-1) & \cdots & \sigma(0) & \sigma(1) & \cdots & 0 \\ 0 & \sigma(q) & \cdots & \sigma(1) & \sigma(0) & \cdots & 0 \\ \vdots & \vdots & & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & \sigma(0) \end{bmatrix} .$$

We define

$$(95) \qquad \mathcal{S}_{q,T} = \{\sigma(0),\ldots,\sigma(q) : \underset{\sim}{\Sigma}_T \text{ is positive definite}\} ,$$

$$(96) \qquad \mathcal{S}_{q,\infty} = \{\sigma(0),\ldots,\sigma(q) : \underset{\sim}{\Sigma}_{T'} \text{ is positive definite } \forall T' \geq q+1\} .$$

A vector $\underset{\sim}{\sigma} = [\sigma(0),\ldots,\sigma(q)]' \in \mathcal{S}_{q,\infty}$ if and only if there exist real $\gamma_0,\ldots,\gamma_q$ such that (93) is satisfied. Alternatively, $\underset{\sim}{\sigma} \in \mathcal{S}_{q,\infty}$ if and only if

$$(97) \qquad 2\pi f(\lambda) = \sum_{h=-q}^{q} \sigma(\lambda) e^{i\lambda h} \geq 0 , \quad \forall\lambda .$$

Clearly $\mathcal{S}_{q,q+1} \supset \mathcal{S}_{q,q+2} \supset \cdots \supset \mathcal{S}_{q,\infty}$.

An alternative set of parameters consists of γ_0 and $\theta_1,\dots,\theta_q$, the roots of

(98) $$\gamma_0\,\theta^q + \gamma_1\,\theta^{q-1} + \dots + \gamma_q = 0 \quad .$$

There is no loss in generality in requiring $|\theta_i| \le 1$, $i=1,\dots,q$. [See Section 7.5.2 of Anderson (1971).] A moving average process is <u>invertible</u> if $|\theta_i| < 1$, $i=1,\dots,q$. It is <u>noninvertible</u> if $|\theta_i| = 1$ for at least one value of i.

The derivatives of the loglikelihood function with respect to $\gamma_0,\dots,\gamma_q$ are

(99) $$\left(\frac{\partial \log L}{\partial \gamma_0}, \dots, \frac{\partial \log L}{\partial \gamma_q}\right) = \left(\frac{\partial \log L}{\partial \sigma(0)}, \dots, \frac{\partial \log L}{\partial \sigma(q)}\right) \begin{bmatrix} \frac{\partial \sigma(0)}{\partial \gamma_0} & \cdots & \frac{\partial \sigma(0)}{\partial \gamma_q} \\ \vdots & & \vdots \\ \frac{\partial \sigma(q)}{\partial \gamma_0} & \cdots & \frac{\partial \sigma(q)}{\partial \gamma_q} \end{bmatrix} .$$

Let $\hat{\underset{\sim}{\sigma}} = (\hat{\sigma}(0),\dots,\hat{\sigma}(q))'$ satisfy

(100) $$\underset{\sim}{0} = \left(\frac{\partial \log L}{\partial \sigma(0)}, \dots, \frac{\partial \log L}{\partial \sigma(q)}\right) .$$

If there is no $\hat{\underset{\sim}{\sigma}} \in \mathcal{S}_{q,\infty}$, then we must have $|\underset{\sim}{J}| = 0$ for $\underset{\sim}{\gamma} = \hat{\underset{\sim}{\gamma}}$, where

$$(101) \qquad \underset{\sim}{J} = \left[\frac{\partial\sigma(i)}{\partial\gamma_j}\right]$$

is the Jacobian matrix and $\hat{\underset{\sim}{\gamma}} = (\hat{\gamma}_0,\ldots,\hat{\gamma}_q)$ satisfies (99) set equal to 0. We shall show that in this case $\hat{\underset{\sim}{\gamma}} = (\hat{\gamma}_0,\ldots,\hat{\gamma}_q)$ corresponds to a noninvertible process.

Theorem 6.1. Let $\theta_1,\ldots,\theta_q$ denote the roots of (98). Then

$$(102) \qquad |\underset{\sim}{J}| = 2\gamma_0^{q+1} \prod_{i=1}^{q}(1-\theta_i) \prod_{i=1}^{q}(1+\theta_i) \prod_{i<j}(1-\theta_i\theta_j) .$$

We see that $|\underset{\sim}{J}|$ vanishes if and only if the characteristic polynomial has at least one root of absolute value 1. Hence we have

Corollary 6.1. The Jacobian $|\underset{\sim}{J}|$ vanishes at $\underset{\sim}{\gamma} = (\gamma_0,\ldots,\gamma_q)$ if and only if the corresponding process (91) is noninvertible.

Proof of Theorem 6.1. The Jacobian is

$$(103) \quad |\underset{\sim}{J}| = \begin{vmatrix} \frac{\partial\sigma(0)}{\partial\gamma_0} & \frac{\partial\sigma(0)}{\partial\gamma_1} & \cdots \\ \frac{\partial\sigma(1)}{\partial\gamma_0} & \frac{\partial\sigma(1)}{\partial\gamma_1} & \cdots \\ \vdots & \vdots & \\ \frac{\partial\sigma(q)}{\partial\gamma_0} & \frac{\partial\sigma(q)}{\partial\gamma_1} & \cdots \end{vmatrix} = \begin{vmatrix} 2\gamma_0 & 2\gamma_1 & \cdots & 2\gamma_q \\ \gamma_1 & \gamma_0+\gamma_2 & \cdots & \gamma_{q-1} \\ \gamma_2 & \gamma_3 & \cdots & \gamma_{q-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{q-1} & \gamma_q & \cdots & \gamma_1 \\ \gamma_q & 0 & \cdots & \gamma_0 \end{vmatrix}$$

$$
= 2 \begin{vmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_q \\ \gamma_1 & \gamma_0+\gamma_2 & \cdots & \gamma_{q-1} \\ \vdots & \vdots & & \vdots \\ \gamma_{q-1} & \gamma_q & \cdots & \gamma_1 \\ \gamma_q & 0 & \cdots & \gamma_0 \end{vmatrix}
$$

$$
= 2\gamma_0^{q+1} \begin{vmatrix} 1 & \alpha_1 & \cdots & \alpha_q \\ \alpha_1 & 1+\alpha_2 & \cdots & \alpha_{q-1} \\ \vdots & \vdots & & \vdots \\ \alpha_{q-1} & \alpha_q & \cdots & \alpha_1 \\ \alpha_q & 0 & \cdots & 1 \end{vmatrix} ,
$$

where $\alpha_i = \gamma_i/\gamma_0$, $i = 0,1,\ldots,q$. We can write

(104) $$1+\alpha_1 x + \ldots + \alpha_q x^q = (1-\theta_1 x) \ldots (1-\theta_q x) ;$$

that is, $\alpha_1 = -\Sigma\theta_i$, $\alpha_2 = \sum_{i<j} \theta_i\theta_j$, etc. Then the last determinant becomes a polynomial in the θ_i's when the α_i's are expressed in terms of θ_i's .

Now we multiply row i by $\theta_1^i + \theta_1^{-i}$ and add to row 0 , $i = 1,\ldots,q$. Then row 0 becomes

(105) $$\left(\sum_{i=0}^{q} \alpha_i\theta_1^i + \sum_{i=0}^{q} \alpha_i\theta_1^{-i} ,\ \theta_1^{-1} \sum_{i=0}^{q} \alpha_i\theta_1^i + \theta_1 \sum_{i=0}^{q} \alpha_i\theta_1^{-i} ,\right.$$
$$\left. \ldots ,\ \theta_1^{-q} \sum_{i=0}^{q} \alpha_i\theta_1^i + \theta_1^q \sum_{i=0}^{q} \alpha_i\theta_1^{-i} \right) .$$

This follows from the evaluation of the j-th column:

(106) $$\alpha_j(\theta_1^0+\theta_1^{-0}) + (\alpha_{j-1}+\alpha_{j+1})(\theta_1^1+\theta_1^{-1}) + (\alpha_{j-2}+\alpha_{j+2})(\theta_1^2+\theta_1^{-2}) + \ldots$$

$$= \sum_{i=0}^{q} \alpha_i(\theta_1^{-j+i}+\theta_1^{j-i}) = \theta_1^{-j}\sum_{i=0}^{q}\alpha_i\theta_1^i + \theta_1^j\sum_{i=0}^{q}\alpha_i\theta_1^{-i} .$$

However $\Sigma_{i=0}^q \alpha_i\theta_1^{-i} = \theta_1^{-q}\sum_{i=0}^{q}\alpha_i\theta_1^{q-i} = 0$ because θ_1 is a root of the characteristic polynomial. Hence, the first row becomes

(107) $$\left(\sum_{i=0}^{q}\alpha_i\theta_1^i ,\ \theta_1^{-1}\sum_{i=0}^{q}\alpha_i\theta_1^i , \ldots, \theta_1^{-q}\sum_{i=0}^{q}\alpha_i\theta_1^i\right) .$$

Hence

(108) $$\sum_{i=0}^{q}\alpha_i\theta_1^i = (1-\theta_1^2)(1-\theta_1\theta_2)\ldots(1-\theta_1\theta_q)$$

is a factor of the determinant.

The above operation can be done with the other θ_i's. Hence we see that $\Pi_{i\leq j}(1-\theta_i\theta_j)$ is a factor of the determinant

(109) $$\begin{vmatrix} 1 & \alpha_1 & \ldots & \alpha_q \\ \alpha_1 & 1+\alpha_2 & \ldots & \alpha_{q-1} \\ \vdots & \vdots & & \vdots \\ \alpha_q & 0 & \ldots & 1 \end{vmatrix} = \begin{vmatrix} 1 & \ldots & (-1)^q\Pi\theta_i \\ -\Sigma\theta_i & & \vdots \\ \vdots & & -\Sigma\theta_i \\ (-1)^q\Pi\theta_i & \ldots & 1 \end{vmatrix} .$$

Now consider the degree of this polynomial in the θ_i's . The highest term comes from

(110) $$\pm\,\alpha_q^{q+1} = \pm\,\Pi\,\theta_i^{q+1} .$$

The highest degree term in $\Pi_{i\le j}(1-\theta_i\theta_j)$ is $\pm\Pi_{i\le j}\theta_i\theta_j = \pm\Pi_i\theta_i^{q+1}$. These agree. Hence $|\underset{\sim}{J}|$ is $c\gamma_0^{q+1}\Pi_{i\le j}(1-\theta_i\theta_j)$ for some constant c. Considering the constant term we obtain $c=2$. Hence

$$|\underset{\sim}{J}| = 2\gamma_0^{q+1} \prod_{i\le j}(1-\theta_i\theta_j) \tag{111}$$

$$= 2\gamma_0^{q+1}\,\Pi(1-\theta_i)\,\Pi(1+\theta_i)\prod_{i<j}(1-\theta_i\theta_j)\,. \qquad \text{Q.E.D.}$$

Now let us find the null vector of the Jacobian matrix $\underset{\sim}{J}$ when it is singular. We have shown that $|\underset{\sim}{J}| = 0$ if and only if $(\gamma_0,\ldots,\gamma_q)$ corresponds to a noninvertible process. The proof gives an explicit expression of a null vector $\underset{\sim}{\eta}$ such that

$$\underset{\sim}{\eta}'\,\underset{\sim}{J} = \underset{\sim}{0}'\,. \tag{112}$$

Suppose that $\underset{\sim}{\gamma} = (\gamma_0,\ldots,\gamma_q)'$ corresponds to a noninvertible process. Then there exists a frequency $\nu(0\le\nu\le\pi)$ such that the spectral density is zero at ν; that is

$$2\pi f(\nu) = \left|\sum_{j=0}^{q}\gamma_j\,e^{i\nu j}\right|^2 \tag{113}$$

$$= \sigma(0)+2\sigma(1)\cos\nu+\ldots+2\sigma(q)\cos\nu q$$

$$= 0\,.$$

It follows that $\Sigma_{j=0}^{q}\gamma_j\,e^{i\nu j}=0$; and $\Sigma_{j=0}^{q}\gamma_j\,e^{-i\nu j}=0$. Let $\theta_1 = e^{i\nu}$ in the proof. Then $\theta_1^j+\theta_1^{-j} = 2\cos j\nu$. Let

$$(114) \qquad \underset{\sim}{\eta}' = (1, \theta_1 + \theta_1^{-1}, \ldots, \theta_1^q + \theta_1^{-q})$$
$$= (1, 2\cos\nu, \ldots, 2\cos q\nu) \quad .$$

Then the above proof shows that

$$(115) \qquad (2, 2\cos\nu, \ldots, 2\cos q\nu) \begin{pmatrix} \gamma_0, \gamma_1, \cdots \gamma_q \\ \gamma_1 & \gamma_{q-1} \\ \vdots & \vdots \\ \gamma_q & \gamma_0 \end{pmatrix}$$

$$= (1, 2\cos\nu, \ldots, 2\cos q\nu) \begin{pmatrix} 2\gamma_0, 2\gamma_1, \ldots, 2\gamma_q \\ \gamma_1 & \gamma_{q-1} \\ \vdots & \vdots \\ \gamma_q & \gamma_0 \end{pmatrix}$$

$$= \underset{\sim}{\eta}' \underset{\sim}{J} = \underset{\sim}{0}' \quad .$$

Therefore we have proved:

Theorem 6.2. Let $\gamma_0, \ldots, \gamma_q$ correspond to a noninvertible process. Let ν be such that the spectral density $f(\nu) = 0$. Then $\underset{\sim}{\eta}' = (1, 2\cos\nu, \ldots, 2\cos q\nu)$ is a null vector (from the left) of the Jacobian matrix $\underset{\sim}{J}$.

This theorem can be given an alternative proof as follows. Let

$$(116) \quad 2\pi f(\nu; a_0, \ldots, a_q) = (a_0^2 + \ldots + a_q^2) + 2(a_0 a_1 + \ldots + a_{q-1} a_q) \cos\nu$$
$$+ \ldots + 2\, a_0 a_q \cos(\nu q) \; .$$

Then $f(\nu;\, a_0,\ldots,a_q) \geq 0$ for all (real) $a_0,\ldots,a_q$. By assumption $f(\nu;\, \gamma_0,\ldots,\gamma_q) = 0$. Let $a_1 = \gamma_1,\ldots,a_q = \gamma_q$ be fixed and consider $f(\nu;\, a_0,\, \gamma_1,\ldots,\gamma_q)$ as a function of a_0. It attains a minimum (=0) at $a_0 = \gamma_0$. Hence

(117) $$\frac{\partial f}{\partial a_0}(\nu;\, \gamma_0,\gamma_1,\ldots,\gamma_q) = 0 \quad .$$

This gives

(118) $$(1,\, 2\cos\nu,\ldots,2\cos q\nu)\begin{pmatrix} 2\gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_q \end{pmatrix} = 0 \; .$$

Similarly considering $\gamma_1,\ldots,\gamma_q$ in sequence we obtain

(119) $$(1,\, 2\cos\nu,\ldots,\, 2\cos q\nu)\, \underset{\sim}{J} = \underset{\sim}{0}' \; .$$

This completes the alternative proof.

Now recall the likelihood equation:

(120) $$\left(\frac{\partial \log L}{\partial\sigma(0)} ,\ldots, \frac{\partial \log L}{\partial\sigma(q)} \right) \underset{\sim}{J} = \underset{\sim}{0}' \; .$$

If we assume that the rank of $\underset{\sim}{J}$ is q, then the null vector is unique up to a multiplicative constant. Hence we have:

Theorem 6.3. Suppose that the likelihood equation $\left(\frac{\partial \log L}{\partial \gamma_0}, \ldots, \frac{\partial \log L}{\partial \gamma_q}\right)$ $= (0,\ldots,0)$ has a noninvertible solution $\gamma_0,\ldots,\gamma_q$ and the rank of $\underset{\sim}{J}$ is q. Then there exists a unique $\nu \ge 0$ such that the spectral density is zero at ν and

$$\left(\frac{\partial \log L}{\partial \sigma(0)}, \ldots, \frac{\partial \log L}{\partial \sigma(g)}\right) = c(1, 2\cos\nu, \ldots, 2\cos q\nu) \tag{121}$$

for some c.

The rank condition of this theorem is a natural one because it corresponds to smoothness of the boundary of the invertibility region. This can be illustrated by considering the MA(2) case. In terms of (ρ_1,ρ_2) the region has a smooth boundary except for the point (ρ_1,ρ_2) $= (0, -\frac{1}{2})$. See Figure 5 of the next section. Consider the Jacobian matrix

$$\underset{\sim}{J} = \begin{pmatrix} 2\gamma_0 & 2\gamma_1 & 2\gamma_2 \\ \gamma_1 & \gamma_0+\gamma_2 & \gamma_1 \\ \gamma_2 & 0 & \gamma_0 \end{pmatrix}. \tag{122}$$

$\underset{\sim}{J}$ is of rank 1 if and only if row 1 is proportional to row 3 and row 2 is proportional to row 3. But this implies $2\gamma_1 = 0$, $\gamma_0+\gamma_2 = 0$, or $\gamma_1 = 0$, $\gamma_2 = -\gamma_0$. (Conversely if $\gamma_1 = 0$, $\gamma_2 = -\gamma_0$, then $\underset{\sim}{J}$ is of rank 1.) Hence this case corresponds to $(\rho_1,\rho_2) = (0, -\frac{1}{2})$. For other boundary points the rank of $\underset{\sim}{J} = 2$.

Geometric interpretation.

The above results can be interpreted from a geometric viewpoint. Consider the set $\mathcal{S}_{q,\infty}$ again. This set is convex. This follows from the fact that $(\sigma(0),\ldots,\sigma(q)) \in \mathcal{S}_{q,\infty}$ if and only if $\sigma(0)+2\sigma(1)\cos\lambda + \ldots + 2\sigma(q)\cos(q\lambda) \geq 0$ for all λ. Let $\underset{\sim}{\sigma}^{(0)} = (\sigma^0,\ldots,\sigma^q) = (\sigma^0(\underset{\sim}{\gamma}),\ldots,\sigma^q(\underset{\sim}{\gamma}))$ be a boundary point of $\mathcal{S}_{q,\infty}$ and let P be a supporting hyperplane at $\underset{\sim}{\sigma}^{(0)}$:

$$(\sigma(0)-\sigma^0)\, c_0 + \ldots + (\sigma(q)-\sigma^q)\, c_q \leq 0 \tag{123}$$

for all $(\sigma(0),\ldots,\sigma(q)) \in \mathcal{S}_{q,\infty}$. See Figure 4. $\underset{\sim}{c}' = (c_0,\ldots,c_q)$ is the

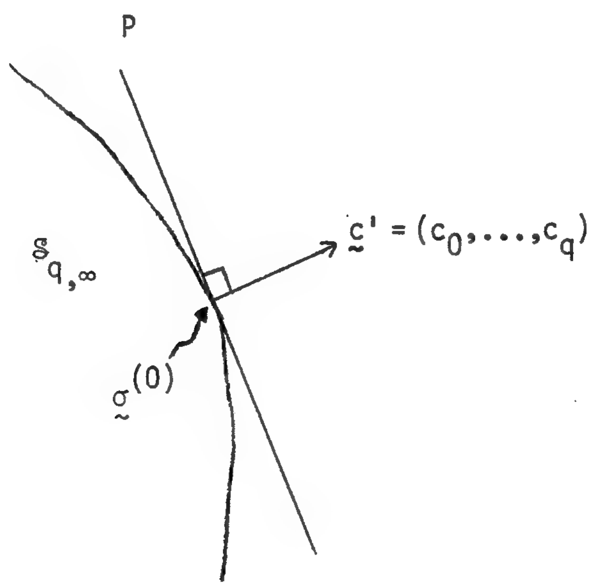

Figure 4.

normal vector to the supporting hyperplane. Actually $\sigma^0 c_0 + \dots + \sigma^q c_q = 0$ because $\mathcal{S}_{q,\infty}$ is a cone. This can be verified by letting $(\sigma(0), \dots, \sigma(q)) = t(\sigma^0, \dots, \sigma^q)$ and considering (123) for $t > 1$ and $0 < t < 1$. Let $\gamma_1, \dots, \gamma_q$ be fixed and let γ_0 be changed by a small amount $\Delta\gamma_0$. Then $\sigma(0) - \sigma^0 = \frac{\partial\sigma(0)}{\partial\gamma_0}\Delta\gamma_0, \dots, \sigma(q) - \sigma^q = \frac{\partial\sigma(q)}{\partial\gamma_0}\Delta\gamma_0$. Substituting this into (123) we have $\Delta\gamma_0 \, \Sigma_{i=0}^q \, \partial\sigma(i)/\partial\gamma_0 \cdot c_i \le 0$. $\Delta\gamma_0$ can be positive or negative, hence $\Sigma_{i=0}^q \, \partial\sigma(i)/\partial\gamma_0 \cdot c_i = 0$. Namely the infinitesimal displacement of $(\sigma(0), \dots, \sigma(q))$ lies in the supporting hyperplane. This consideration can be applied to $\gamma_1, \dots, \gamma_q$ as well. In matrix form we then have

$$\underset{\sim}{c}'\underset{\sim}{J} = \underset{\sim}{0}' . \tag{124}$$

Namely the normal vector to the supporting hyperplane is a null vector. Hence under the same assumptions of the last theorem we have

<u>Theorem 6.4</u>. Let the assumptions of Theorem 6.3 hold. Then there exists a unique supporting hyperplane at the boundary point $\underset{\sim}{\sigma}^0$ and the gradient of $\log L$ with respect to $\sigma(0), \dots, \sigma(q)$ is proportional to the normal vector of the hyperplane.

This theorem implies that if the boundary point $\underset{\sim}{\sigma}$ corresponds to a relative maximum of the likelihood function, then the likelihood in terms of $(\sigma(0), \dots, \sigma(q))$ increases most steeply in the direction orthogonal to $\mathcal{S}_{q,\infty}$.

7. THE MOVING AVERAGE OF ORDER 2

The results of the previous section can be illustrated by considerng the MA(2) process. The region of (ρ_1,ρ_2) corresponding to the MA(2) processes is given in Section 3.4 of Box and Jenkins (1976). It is the intersection of $\mathcal{S}_{2,\infty}$ of the previous section and the plane $\sigma(0)=1$. The boundary is given by $\rho_2=\rho_1-\frac{1}{2}$, $\rho_2=-\rho_1-\frac{1}{2}$, $\rho_1^2+8(\rho_2-\frac{1}{4})^2=\frac{1}{2}$ corresponding to root -1,1, and two complex conjugate roots of absolute value 1, respectively. For $T=3$, $\underset{\sim}{\Sigma}_3$ is positive semidefinite if and only if $-1\le\rho_1\le 1$ and $\rho_1^2\le(1+\rho_2)/2$. For $T=4$, $\underset{\sim}{\Sigma}_4$ is positive semidefinite if and only if

$$(125)\qquad \rho_1^2 \le (\rho_2-1)^2+\frac{1}{2}-[\{(\rho_2-1)^2+\frac{1}{2}\}^2-(1-\rho_2^2)^2]^{1/2} \quad .$$

For larger T , explicit expression for the positive definiteness of $\underset{\sim}{\Sigma}_T$ seems difficult to obtain. In Figure 5 boundaries corresponding to $T = 3,4,5,6,\infty$ are plotted. For $T = 5,6$ the boundaries are computed numerically. We see that for finite T the regions are strictly larger than the region arising from MA(2) processes $(T=\infty)$. For $\rho_2 = 0$, the boundary points are the same as for MA(1) case, namely $\rho_1 = \pm 1/2 \cos(\frac{\pi}{T+1})$. For $\rho_1 = 0$, the explicit expressions of limits can be obtained as follows. If $\rho_1=0$, then $y_1,y_3,y_5,\ldots$ (odd indices) and $y_2,y_4,\ldots$ (even indices) are uncorrelated and each subseries forms a MA(1) process with parameter ρ_2 . Hence $\underset{\sim}{\Sigma}_T$ is positive semidefinite if and only if submatrices for even indices and odd indices are both positive semidefinite. It follows that the limits are given by

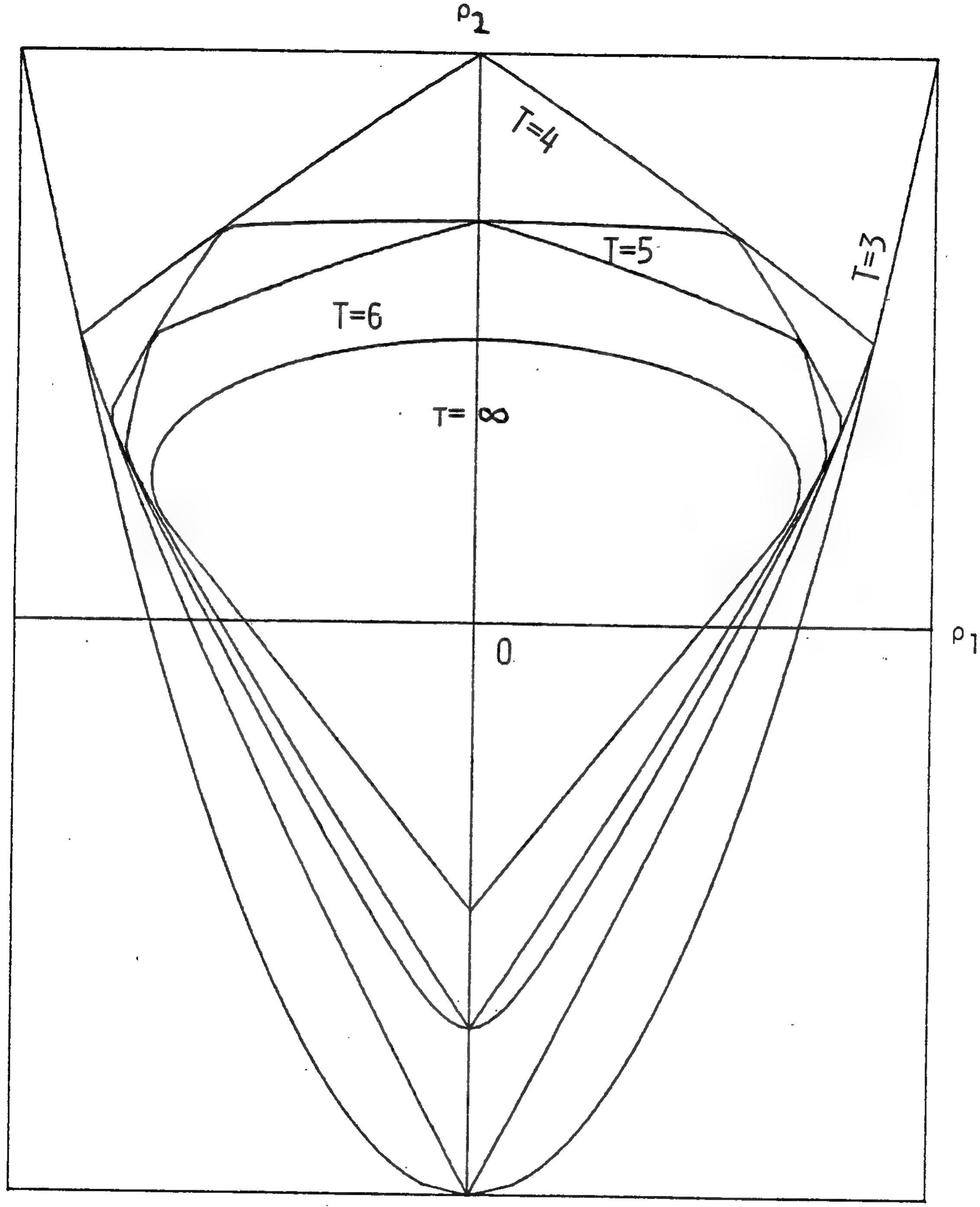

Figure 5.

(126) $$\rho_2 = \pm \frac{1}{2\cos\frac{\pi}{[(T+1)/2]+1}} ,$$

where $[(T+1)/2]$ equals $T/2$ if T is even and $(T+1)/2$ if T is odd.

Detailed analysis of full likelihood function of the MA(2) process seems to be difficult to carry out.

8. THE AUTOREGRESSIVE MOVING AVERAGE MODEL

The ARMA(p,q) model is

(127) $$\sum_{j=0}^{p} \beta_j\, y_{t-j} = \sum_{j=0}^{q} \alpha_j\, v_{t-j} ,$$

where $\alpha_0 = \beta_0 = 1$, and $\{v_t\}$ is a sequence of independent, identically distributed random variables with $\mathcal{E}v_t = 0$ and $\mathcal{E}v_t^2 = \sigma_v^2$. Let

(128) $$u_t = \sum_{j=0}^{q} \alpha_j v_{t-j} = \sum_{j=0}^{q} \gamma_j w_{t-j} ,$$

when $\gamma_j = \sigma_v \alpha_j$, $j = 0,1,\ldots,q$, and $w_t = v_t/\sigma_v$. The autocovariances of the unobservable $\{u_t\}$ process are

(129) $$\begin{aligned} \sigma_u(h) &= \sum_{j=0}^{q-h} \gamma_j \gamma_{j+h} , & h &= 0,1,\ldots,q , \\ &= 0 , & h &> q . \end{aligned}$$

Let $\underset{\sim}{\beta} = (\beta_1,\ldots,\beta_p)'$, $\underset{\sim}{\sigma}_u = [\sigma_u(0),\ldots,\sigma_u(q)]'$, and $\underset{\sim}{\gamma} = (\gamma_0,\ldots,\gamma_q)'$. Alternative parametrizations are $(\underset{\sim}{\beta},\underset{\sim}{\gamma})$ (or equivalently $\underset{\sim}{\beta}$, $\underset{\sim}{\alpha}$, and σ_v^2) and $(\underset{\sim}{\beta},\underset{\sim}{\sigma}_u)$.

The derivatives of the likelihood function with respect to the components of $\underset{\sim}{\beta}$ and $\underset{\sim}{\gamma}$ are

$$\left(\frac{\partial L}{\partial \underset{\sim}{\beta}'} , \frac{\partial L}{\partial \underset{\sim}{\gamma}'}\right) = \left(\frac{\partial L}{\partial \underset{\sim}{\beta}'} , \frac{\partial L}{\partial \underset{\sim}{\sigma}_u'}\right)\begin{bmatrix} \underset{\sim}{I} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{J}_u \end{bmatrix} , \tag{130}$$

where $\underset{\sim}{J}_u = (\partial\underset{\sim}{\sigma}_u/\partial\underset{\sim}{\gamma}')$. As in the case of MA(q) , if there is a solution of the likelihood equations for $\underset{\sim}{\beta}$ and $\underset{\sim}{\gamma}$ such that the corresponding vector $(\partial L/\partial\underset{\sim}{\beta}' , \partial L/\partial\underset{\sim}{\sigma}_u')$ is not $\underset{\sim}{0}'$, then the matrix on the right-hand side of (130) must be singular, that is, $|\underset{\sim}{J}_u| = 0$ at this vector $\hat{\underset{\sim}{\gamma}}$. This solution is noninvertible. The analysis of the MA(q) model can be carried over to the ARMA(p,q) model.

9. THE AUTOREGRESSIVE MOVING-AVERAGE PROCESS OF ORDER 1 AND 1

The considerations of the previous section can be illustrated by ARMA(1,1) process. Let $y_1,\ldots,y_T$ be T successive observations from an ARMA(1,1) process, that is, the y_t's satisfy

$$y_t + \beta y_{t-1} = v_t + \alpha v_{t-1} = u_t . \tag{131}$$

Let $E(u_t^2) = \sigma_u^2$ and $\rho = \alpha/(1+\alpha^2)$. In terms of the parametrization (β,ρ,σ_u^2) the autocovariances of the process are given by

$$\sigma(0) = \sigma_u^2 \frac{1-2\beta\rho}{1-\beta^2}, \tag{132}$$

$$\sigma(1) = \sigma_u^2 \frac{\rho+\beta^2\rho-\beta}{1-\beta^2},$$

$$\sigma(h+1) = -\beta\sigma(h), \qquad h = 1,2,\ldots,$$

$$\sigma(-h) = \sigma(h), \qquad h = 1,2,\ldots.$$

To investigate the T-dimensional covariance matrix $\underset{\sim}{\Sigma}_T$, it is useful to consider the following transformation:

$$\begin{pmatrix} y_1 \\ u_2 \\ u_3 \\ \vdots \\ u_T \end{pmatrix} = \begin{pmatrix} 1 & & & & \\ \beta & 1 & & & \\ & \beta & 1 & & \\ & & \ddots & \ddots & \\ & & & \beta & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_T \end{pmatrix}. \tag{133}$$

Using $\mathrm{cov}(y_1,u_2) = \mathrm{cov}(u_1 - \beta y_0,u_2) = \mathrm{cov}(u_1,u_2) = \rho\sigma_u^2$, $\mathrm{var}(y_1) = \sigma_u^2(1-2\beta\rho)/(1-\beta^2)$, we obtain

$$\mathrm{Var}\begin{pmatrix} y_1 \\ u_2 \\ u_3 \\ \vdots \\ u_T \end{pmatrix} = \begin{pmatrix} 1 & & & & \\ \beta & 1 & & & \\ & \beta & 1 & & \\ & & \ddots & \ddots & \\ & & & & 1 \end{pmatrix} \underset{\sim}{\Sigma}_T \begin{pmatrix} 1 & \beta & & & \\ & 1 & \beta & & \\ & & 1 & \ddots & \\ & & & \ddots & \ddots \\ & & & & 1 \end{pmatrix} \tag{134}$$

$$= \sigma_u^2 \begin{bmatrix} \frac{1-2\beta\rho}{1-\beta^2} & \rho & & & \\ \rho & 1 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & & 1 & \rho \\ & & & \rho & 1 \end{bmatrix} .$$

Hence $\underset{\sim}{\Sigma}_T$ is positive semidefinite if and only if the matrix on the right-hand side of (134) is positive semidefinite. Now define the determinant of the $T \times T$ matrix

$$D_T = \begin{vmatrix} 1 & \rho & & & \\ \rho & 1 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & & 1 & \rho \\ & & & \rho & 1 \end{vmatrix} . \tag{135}$$

Then the determinant of (134) is σ_u^2 times

$$\begin{vmatrix} \frac{1-2\beta\rho}{1-\beta^2} & \rho & & & \\ \rho & 1 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & & 1 & \rho \\ & & & \rho & 1 \end{vmatrix} = \frac{1}{1-\beta^2} \begin{vmatrix} 1-2\beta\rho & (1-\beta^2)\rho & & & \\ \rho & 1 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & & 1 & \rho \\ & & & \rho & 1 \end{vmatrix} \tag{136}$$

$$= \frac{1}{1-\beta^2} (D_T - 2\beta\rho D_{T-1} + \beta^2\rho^2 D_{T-2}) .$$

From Lemma 6.7.9 of Anderson (1971) we have

$$D_T = \frac{1}{\sqrt{1-4\rho^2}} \left[\left(\frac{1+\sqrt{1-4\rho^2}}{2} \right)^{T+1} - \left(\frac{1-\sqrt{1-4\rho^2}}{2} \right)^{T+1} \right] . \tag{137}$$

Here we are interested in the case $|\rho| \geq \frac{1}{2}$, because if $|\rho| < \frac{1}{2}$ then the covariance matrix is clearly positive definite. Let $\rho \geq \frac{1}{2}$ without loss of generality. Now

$$\frac{1\pm\sqrt{1-4\rho^2}}{2} = \frac{1\pm i\sqrt{4\rho^2-1}}{2} = \rho\,\frac{1\pm i\sqrt{4\rho^2-1}}{2\rho} = \rho(\cos\theta \pm i \sin\theta) = \rho e^{\pm i\theta}, \tag{138}$$

where $\theta = \tan^{-1}\sqrt{4\rho^2-1}$. Then

$$\left(\frac{1\pm\sqrt{1-4\rho^2}}{2} \right)^k = \rho^k(\cos k\theta \pm i \sin k\theta) . \tag{139}$$

Therefore

$$\begin{aligned} &D_T - 2\beta\rho\, D_{T-1} + \beta^2\rho^2\, D_{T-2} \\ &= \frac{1}{i\sqrt{4\rho^2-1}} [\rho^{T+1}\, 2\, i \sin(T+1)\theta - 2\beta\rho\rho^T 2\, i \sin T\theta + \beta^2\rho^2\rho^{T-1}\, 2\, i \sin(T-1)\theta] \\ &= \frac{2\rho^{T+1}}{\sqrt{4\rho^2-1}} [\sin(T+1)\theta - 2\beta \sin T\theta + \beta^2 \sin(T-1)\theta] . \end{aligned} \tag{140}$$

Hence the determinant is zero if and only if

$$\beta^2 \sin(T-1)\theta - 2\beta \sin T\theta + \sin(T+1)\theta = 0 . \tag{141}$$

Solving (140) for β we obtain

(142) $$\beta = \frac{\sin T\theta \pm \sin\theta}{\sin(T-1)\theta} .$$

Considering $\theta \doteq 0$ ($\rho \doteq 1/2$) we have $\beta \doteq (T\theta \pm \theta)/(T-1)\theta$. This has to be less than 1. Hence we take the negative sign. So

(143) $$\beta = \frac{\sin T\theta - \sin\theta}{\sin(T-1)\theta} .$$

Using $\sin T\theta - \sin\theta = 2\sin\frac{T-1}{2}\theta\cos\frac{T+1}{2}\theta$, $\sin(T-1)\theta = 2\sin\frac{T-1}{2}\theta\cos\frac{T-1}{2}\theta$, we can write

(144) $$\beta = \cos\frac{T+1}{2}\theta \Big/ \cos\frac{T-1}{2}\theta .$$

Equations (143) and (144) give the boundaries of the region $\underset{\sim}{\Sigma}_T$: positive semidefinite in (ρ,β)-plane. For $\beta = 1$ we have $\theta = 0$ or $\rho = \frac{1}{2}$. For $\beta = 0$ we have MA(1) process and we obtain $\rho = 1/2\cos\frac{\pi}{T+1}$. This can be verified by setting $\cos\frac{T+1}{2}\theta = 0$. For $\beta = -1$ we have $\cos\frac{T+1}{2}\theta = -\cos\frac{T-1}{2}\theta$ or $\frac{T}{2}\theta = \frac{\pi}{2}$. Hence $\theta = \frac{\pi}{T}$ or $\rho = 1/2\cos(\pi/T)$. A plot of the boundaries is given in Figure 6.

For $\rho \le 0$ the regions are symmetric about the origin. Again we see that for finite T the region is strictly larger than the region corresponding to ARMA(1,1) processes $(T = \infty)$.

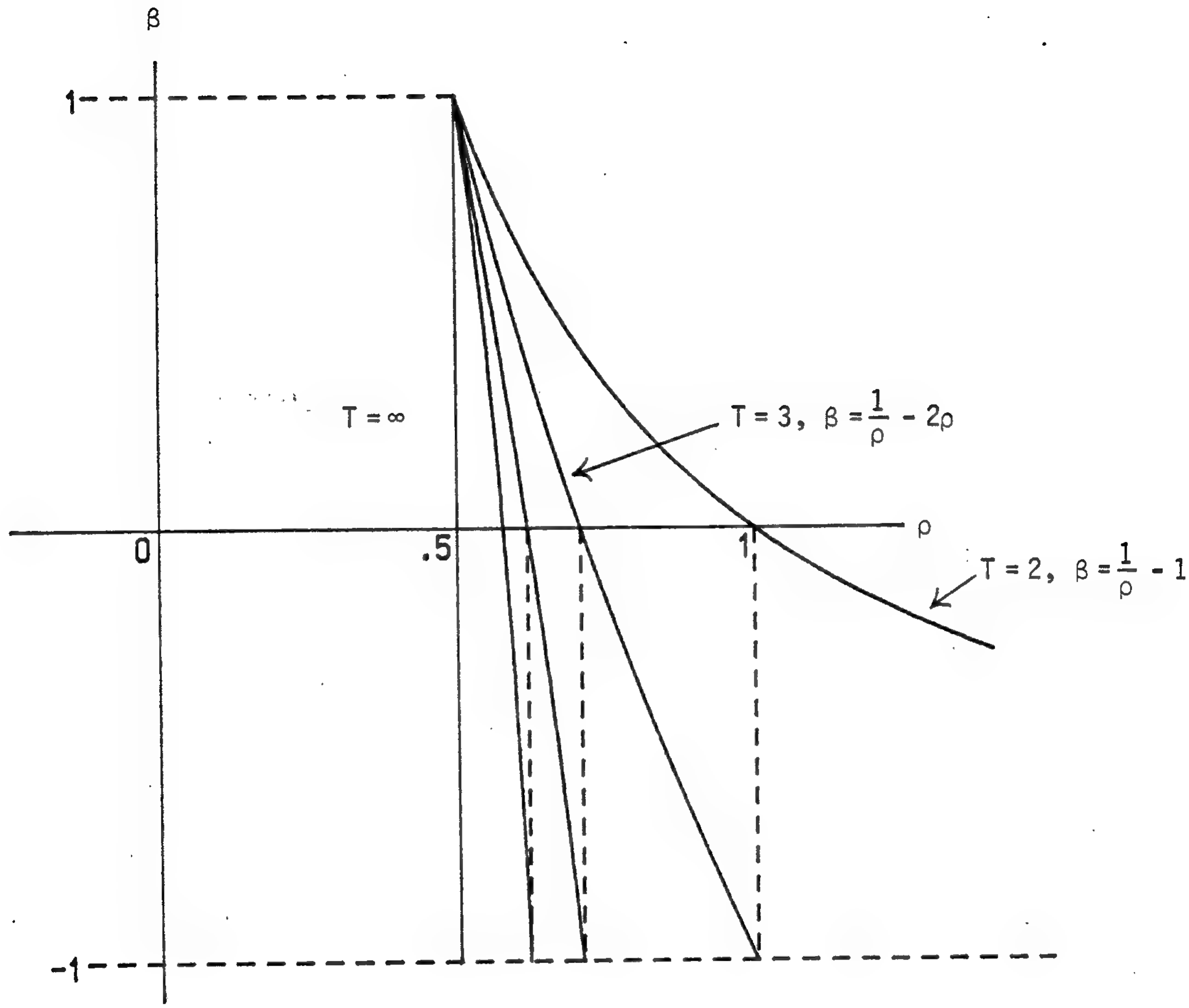

Figure 6.

Appendix A

A1
$$\Sigma \frac{d_t}{1+d_t} = \frac{1}{2}\frac{d}{d\rho}\log|\underset{\sim}{R}|\Big|_{\rho=1/2} .$$

Since $\log|\underset{\sim}{R}|$ is continuously differentiable in ρ we can let ρ approach 1/2 from above. Then as in Section 9

$$|\underset{\sim}{R}| = D_T = \rho^T \frac{\sin(T+1)\theta}{\sin\theta}$$

$$= \rho^T \frac{\sin(T\theta)\cos\theta + \cos(T\theta)\sin\theta}{\sin\theta}$$

$$= \rho\cos\theta\, D_{T-1} + \rho^T\cos(T\theta)$$

$$= \frac{1}{2} D_{T-1} + \rho^T\cos(T\theta) ,$$

where $\theta = \tan^{-1}[(4\rho^2 - 1)^{1/2}]$ and $\cos\theta = 1/2\rho$. Hence

$$\frac{d}{d\rho} D_T = \frac{1}{2}\frac{d}{d\rho} D_{T-1} + T\rho^{T-1}\cos(T\theta) - T\rho^T\sin(T\theta)\cdot\frac{d\theta}{d\rho} .$$

Now

$$\frac{d\theta}{d\rho} = \frac{1}{1+4\rho^2-1}\cdot\frac{8\rho}{2(4\rho^2-1)^{1/2}} = \frac{1}{2\rho^2}\frac{2\rho}{(4\rho^2-1)^{1/2}} = \frac{1}{2\rho^2\sin\theta} .$$

Letting $\rho \to 1/2$ (hence $\theta \to 0$) we have

$$\frac{d}{d\rho} D_T\Big|_{\rho=1/2} = \frac{1}{2}\frac{d}{d\rho} D_{T-1}\Big|_{\rho=1/2} - T(T-1)2^{-T+1} ,$$

or

$$a_T = a_{T-1} - 2\,T(T-1) ,$$

where $a_T = 2^T (d/d\rho) D_T \big|_{\rho=1/2}$.

Hence

$$a_T = -2 \sum_{t=1}^{T} t(t-1) = -(2/3)(T+1)T(T-1) .$$

Also

$$D_T \big|_{\rho=1/2} = 2^{-T}(T+1) .$$

Combining these we obtain

$$\sum_{t=1}^{T} \frac{d_t}{1+d_t} = -\frac{T(T-1)}{3} .$$

<u>A2</u> We have to check that for T sufficiently large the left-hand side of (53) has at least k terms. Now $w_t > 0$ is equivalent to

$$d_t < -\frac{T-1}{T+2} = -\frac{1-\frac{1}{T}}{1+\frac{2}{T}}$$

$$= -(1-\tfrac{1}{T})(1-\tfrac{2}{T}+\ldots)$$

$$= -1+\tfrac{3}{T}+0(\tfrac{1}{T^2}) .$$

However,

$$d_{T-j} = -1 + \frac{\pi^2(j+1)^2}{2(T+1)^2} + \dots \ .$$

Hence we see that the left-hand side contains about $\sqrt{T}$ terms.

References

Anderson, T. W. (1971). The Statistical Analysis of Time Series. John Wiley and Sons: New York.

Anderson, T. W., and D. A. Darling (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. Ann. Math. Statist., 23, 193-212.

Anderson, T. W., and Raul P. Mentz (1980). On the structure of the likelihood function of autoregressive and moving average models. Journal of Time Series Analysis, 1, 83-94.

Box, G.E.P., and G. M. Jenkins (1976). Time Series Analysis: Forecasting and Control. Holden-Day: San Francisco.

Cryer, J. D., and J. Ledolter (1981). Small-sample properties of the maximum likelihood estimator in the first-order moving average model. Biometrika, 68, 691-694.

Davidson, J. (1981a). Problems with the estimation of moving average processes. Journal of Econometrics, 16, 295-310.

Davidson, J. (1981b). Small sample properties of estimators of the moving average process. Proceedings of the Econometric Society European Meeting, 1979 (E. G. Charatsis, ed.), 27-62.

Kang, K. M. (1975). A comparison of estimators for moving average processes. Unpublished, Australian Bureau of Statistics.

Pesaran, M. H. (1983). A note on the maximum likelihood estimation of regression models with first order moving average errors with roots on the unit circle. Australian Journal of Statistics, 25, 442-448.

Sargan, J. D. and A. Bhargava (1983). Maximum likelihood estimation of regression models with first order moving average errors when the root lies on the unit circle. Econometrica, 51, 799-820.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 13	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (*and Subtitle*) Why Do Noninvertible Estimated Moving Averages Occur?		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) T. W. Anderson and Akimichi Takemura		8. CONTRACT OR GRANT NUMBER(s) DAAG 29-82-K-0156
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics - Sequoia Hall, Stanford University, Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P-19065-M
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office, Post Office Box 12211, Research Triangle Park, NC 27709		12. REPORT DATE November 1984
		13. NUMBER OF PAGES 55 pp.
14. MONITORING AGENCY NAME & ADDRESS(*if different from Controlling Office*)		15. SECURITY CLASS. (*of this report*) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (*of this Report*)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (*of the abstract entered in Block 20, if different from Report*)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (*Continue on reverse side if necessary and identify by block number*)

Moving average models, maximum likelihood estimation, noninvertible moving average, autoregressive moving average processes.

20. ABSTRACT (*Continue on reverse side if necessary and identify by block number*)

The positive probability that an estimated moving average process is noninvertible is studied for maximum likelihood estimation of a univariate process. Upper and lower bounds for the probability in the first-order case are obtained as well as limits when the sample size tends to infinity. Higher order moving average models and autoregressive moving average models are also treated.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)